All-Possible-Couplings Approach to Measuring Probabilistic Context

Ehtibar N. Dzhafarov
Department of Psychological Sciences, Purdue University, USA (ehtibar@purdue.edu)

Janne V. Kujala
Department of Mathematical Information Technology, University of Jyväskylä, Finland (jvk@iki.fi)

From behavioral sciences to biology to quantum mechanics, one encounters situations where (i) a system outputs several random variables in response to several inputs, (ii) for each of these responses only some of the inputs may "directly" influence it, but (iii) other inputs provide a "context" for this response by influencing its probabilistic relations to other responses. These contextual influences are very different, say, in the classical kinetic theory and in the entanglement paradigm of quantum mechanics, traditionally being interpreted as representing different forms of physical determinism. We show how one can quantify and classify all logically possible contextual influences, based on probabilistic coupling: studying sets of joint distributions imposed on random outputs recorded at different (mutually incompatible) values of inputs.

Keywords: Bell/CHSH inequalities; context; coupling; determinism; EPR paradigm; selective influences.

1. INTRODUCTION

Consider a system with two inputs, α, β, and two random outputs, A, B, about which it is assumed that A is *not* influenced by β, nor B by α. A necessary condition for this *selectivity of influences* assumption is *marginal selectivity* [*1*]: changes in the value of β do not influence the distribution of A, and analogously for α and B. In physics this condition is known under a variety of other names [*2*]. The assumption of selective influences, however, is stronger. It requires the representability of two outputs as

$$A = f(R, \alpha), B = g(R, \beta), \tag{1.1}$$

where f, g are some functions and R a source of randomness that does not depend on α, β [*3-9*]. Such a representability may or may not exist when marginal selectivity is satisfied. According as which is the case, one can think of α and β as being involved in different kinds of *probabilistic context* for the dependence of, respectively, B on β and A on α. Equation (1.1) is traditionally viewed as indicating lack of contextual influences (in physics, classical determinism), but one can still think of various special cases of this "no-context" contextuality. Analogously, one can think of various kinds of violations of (1.1).

We propose a principled way of quantifying and classifying conceivable contextual influences, whether within or outside the scope of (1.1). Our approach is neutral with respect to such issues as causality or what distinguishes direct influences from contextual. We merely accept as a given a diagram of input-output correspondences (e.g., $A \leftarrow \alpha, B \leftarrow \beta$) and study the joint distribution of the outputs at all possible values of the inputs. The interpretation of the diagram is irrelevant insofar as it is compatible with the observed pattern of marginal selectivity. Our approach is maximally general in the sense of applying to arbitrary sets of inputs and outputs (see Appendix A). To demonstrate it by detailed computations, however, we confine ourselves here to binary α, β influencing binary A, B, and even more narrowly, to the "homogeneous" case with the two values of both A and B equiprobable at all values of the inputs. Marginal selectivity then is satisfied trivially with respect to any diagram of input-output correspondences, as all outcomes have probability $1/2$ irrespective of α, β.

The example focal for this paper is Bohm's version of the Einstein-Podolsky-Rosen paradigm (EPR/B) [*10*]: a quantum mechanical (QM) system consisting (in the simplest case) of two entangled spin-$1/2$ particles separated by a spacelike interval. The two inputs here are spin measurements on these particles: input α has two values corresponding to spin axes α_1, α_2 chosen for one particle, and input β has two values cor-

responding to spin axes β_1, β_2 for another particle. Note that the spin measurements on a given particle along two different axes are *noncommuting*, and this is the reason α_1, α_2 can be viewed as (mutually exclusive) values of a single input, and analogously for β_1, β_2 [*11,12*]. The two outputs are spin values recorded: having chosen axes α_i and β_j, $i, j \in \{1, 2\}$, one records A_{ij} for the first particle and B_{ij} for the second, each being a random variable with values "spin-up" and "spin-down." We consider the case when these two outcomes are equiprobable for both A_{ij} and B_{ij}: $\Pr[A_{ij} = up] = \Pr[B_{ij} = up] = 1/2$ for all $i, j \in \{1, 2\}$.

Formally equivalent situations are abundant in behavioral and social sciences [*9,13-17*], where the issue of selective influences was initially introduced in [*18,19*], in the context of information processing architectures. An example of a system here (from our laboratory) can be a human observer who adjusts a visual stimulus until it matches in appearance another visual stimulus, characterized by two properties, α and β (e.g., amplitudes of two Fourier-components), each varying on two levels, α_1, α_2 and β_1, β_2. Denoting by S^1_{ij} and S^2_{ij} the corresponding properties (amplitudes) of the adjusted stimulus in response to α_i, β_j, we define a binary random output A_{ij} as having the value "high" or "low" according as the variable S^1_{ij} is above or below the median of its distribution; output B_{ij} is defined from S^2_{ij} analogously. Marginal selectivity is ensured by construction: $\Pr[A_{ij} = high] = \Pr[B_{ij} = high] = 1/2$ for all $i, j \in \{1, 2\}$.

Another example is a population of people each of whom takes two exams (say, in French and in chemistry) and gets numerical scores S^1 and S^2. Let each score be observed after one of two preparation times (α_1 = "short," α_2 = "long" for S^1, and analogously for S^2). We define a binary random variable A_{ij} as having the value "pass" or "fail", according as the score S^1_{ij} is above or below the median of all scores for α_i, β_j; output B_{ij} is defined analogously.

In an example from a biological domain S^1_{ij} and S^2_{ij} could be activity levels of two neurons tuned to two stimulus properties, α and β, respectively. Making α and β vary on two levels each and defining A_{ij}, B_{ij} with respect to the medians of S^1_{ij}, S^2_{ij} by the same rule as above, we get precisely the same mathematical formulation.

2. FORMS OF CONTEXT (DETERMINISM)

In the following, symbols i, j, k (possibly with primes) always take on values $1, 2$ each. Equation (1.1) is equivalent to the existence of a jointly distributed system

$$\left(H^1_1, H^1_2, H^2_1, H^2_2\right),$$

such that every observable pair A_{ij}, B_{ij} is distributed as H^1_i, H^2_j; in symbols,

$$\left(H^1_i, H^2_j\right) \sim \left(A_{ij}, B_{ij}\right).$$

This statement is known as (a special case of) the *Joint Distribution Criterion* (JDC) [*7,8,14,20,21*]. Its link to (1.1) is given by $H^1_i = f(R, \alpha_i)$, $H^2_j = g(R, \beta_j)$. JDC is a deep criterion that provides a probabilistic foundation for our understanding of the classical (non)contextuality (or determinism) in physics.

Denoting by a_1, a_2 (and b_1, b_2) the two possible values for A_{ij} (resp., B_{ij}), let

$$p_{ij} = \Pr[A_{ij} = a_1, B_{ij} = b_1].$$

JDC in this case is equivalent to four double-inequalities

$$\frac{1 - C}{2} \leq p_{ij} + p_{ij'} + p_{i'j'} - p_{i'j} \leq \frac{1 + C}{2}, \quad (2.1)$$

with $i \neq i'$, $j \neq j'$ and $C = 1$. [*7,8*]. They are known as *Bell/CHSH inequalities* (in the homogeneous form), CHSH acronymizing the authors of [*5*]. The QM theory of the EPR/B paradigm predicts and experimental data confirm violations of the Bell/CHSH inequalities [*22,23*], but QM imposes its own constraint [*5,6*]: inequality (2.1) with $C = \sqrt{2}$. However, there is no analogue or extension of JDC equivalent to this constraint. In other (e.g., behavioral) applications, one cannot exclude a priori the possibility of C exceeding $\sqrt{2}$, falling between 1 and $\sqrt{2}$, or being constrained by some value below 1. Moreover, one can conceive of

$$m \leq p_{ij} + p_{ij'} + p_{i'j'} - p_{i'j} \leq M$$

with $M + m \neq 1$. It would be unsatisfactory if all these possibilities, whether or not empirically realizable, could not be treated within a unified probabilistic framework including JDC as a special case. We construct such a framework, based on the classical (Kolmogorov's) theory of probability and the probabilistic coupling theory [*25*].

3. CONNECTIONS

It is easy to see that for any observable $p = (p_{11}, p_{12}, p_{21}, p_{22})$ one can find a jointly distributed system of binary variables

$$H = \left(H^1_{11}, H^2_{11}, H^1_{12}, H^2_{12}, H^1_{21}, H^2_{21}, H^1_{22}, H^2_{22}\right)$$

such that

$$\left(H^1_{ij}, H^2_{ij}\right) \sim (A_{ij}, B_{ij})$$

for all i, j. The JDC then amounts to additionally assuming that

$$\Pr\left[H^1_{i1} \neq H^1_{i2}\right] = 0, \Pr\left[H^2_{1j} \neq H^2_{2j}\right] = 0,$$

and this is the assumption that is rejected by QM in the EPR/B paradigm. This assumption, however, is not the only way of thinking of H. Since A_{i1} and A_{i2} (or B_{1j} and B_{2j}) occur under mutually exclusive conditions, there is no privileged pairing scheme for their realizations, whence zero values for $\Pr\left[H^1_{i1} \neq H^1_{i2}\right], \Pr\left[H^2_{1j} \neq H^2_{2j}\right]$ are as acceptable a priori as any other. Our approach consists in replacing this assumption with

$$\begin{array}{r}\Pr\left[H^1_{i1} \neq H^1_{i2}\right] = 2\varepsilon^1_i \in [0,1]\,,\\ \Pr\left[H^2_{1j} \neq H^2_{2j}\right] = 2\varepsilon^1_j \in [0,1]\,,\end{array}$$

and characterizing the dependence of (A, B) on (α, β) by properties of the set of all 4-vectors $\varepsilon = \left(\varepsilon^1_1, \varepsilon^1_2, \varepsilon^2_1, \varepsilon^2_2\right)$ that are compatible with or imply certain constraints imposed on the observable vectors $p = (p_{11}, p_{12}, p_{21}, p_{22})$. Having adopted a particular diagram of input-output correspondences (in our case, $A \leftarrow \alpha, B \leftarrow \beta$), we can also say that these sets of ε characterize the contextual role of α, β for B and A, respectively. We call ε a *connection vector*.

To distinguish our approach from other forms and meanings of probabilistic contextualism, e.g., [*26,27,28*], we dub it the "all-possible-couplings" approach. The term "coupling" refers to imposing a joint distribution (say, that of H^1_{11}, H^1_{12}) on random variables that otherwise are not jointly distributed (A_{11} and A_{12}). Although here we do not need to introduce it explicitly, in more complex settings the set of couplings is a construct separate from A and H (see Appendix A).

4. EXTENDED LINEAR FEASIBILITY POLYTOPE (ELFP)

We need this notion to investigate constraints imposed on the observable p. ELFP is the set of all possible (p, ε) for which there exists jointly distributed H^k_{ij} such that, for all i, j,

$$\left[\begin{array}{c}\left(H^1_{ij}, H^2_{ij}\right) \sim (A_{ij}, B_{ij})\\ \text{i.e.}\\ \Pr\left[H^1_{ij} = a_1, H^2_{ij} = b_1\right] = p_{ij}\end{array}\right],$$

$$\left[\begin{array}{c}\left(H^1_{i1}, H^1_{i2}\right) \sim (A_{i1}, A_{i2})\\ \text{i.e.}\\ \Pr\left[H^1_{i1} = a_1, H^1_{i2} = a_2\right] = \varepsilon^1_i\end{array}\right],$$

$$\left[\begin{array}{c}\left(H^2_{1j}, H^2_{2j}\right) \sim (B_{1j}, B_{2j})\\ \text{i.e.}\\ \Pr\left[H^2_{1j} = b_1, H^2_{2j} = b_2\right] = \varepsilon^2_j\end{array}\right].$$

The existence of such an H means the existence of a probability vector Q consisting of the 2^8 joint probabilities

$$\Pr\left[H^1_{11} = h^1_{11}, H^2_{11} = h^2_{11}, \ldots, H^2_{22} = h^2_{22}\right],$$

$h^1_{ij} \in \{a_1, a_2\}$, $h^2_{ij} \in \{b_1, b_2\}$. Let P denote the 2^5-component vector consisting of 2^4 observable probabilities

$$\Pr\left[A_{ij} = a_{ij}, B_{ij} = b_{ij}\right]$$

and 2^4 connection probabilities

$$\Pr\left[A_{i1} = a_{i1}, A_{i2} = a_{i2}\right],\ \Pr\left[B_{1j} = b_{1j}, B_{2j} = b_{2j}\right],$$

$a_{ij} \in \{a_1, a_2\}$, $b_{ij} \in \{b_1, b_2\}$.

Define a $2^5 \times 2^8$ Boolean matrix M whose rows are enumerated in accordance with components of P (i.e., by equalities $[A_{ij} = a_{ij}, B_{ij} = b_{ij}]$, $[A_{i1} = a_{i1}, A_{i2} = a_{i2}]$, or $[B_{1j} = b_{1j}, B_{2j} = b_{2j}]$) and columns in accordance with components of Q (i.e., by equalities $\left[H^1_{11} = h^1_{11}, H^2_{11} = h^2_{11}, \ldots, H^2_{22} = h^2_{22}\right]$). An entry of M contains 1 if and only if the corresponding random variables in the enumerations of its row and its column have the same values: e.g., if a row is enumerated by $[B_{12} = b_{12}, B_{22} = b_{22}]$ and a column by $\left[H^1_{11} = h^1_{11}, \ldots, H^2_{12} = h^2_{12}, \ldots, H^2_{22} = h^2_{22}\right]$, then their intersection contains 1 if and only if $h^2_{12} = b_{12}, h^2_{22} = b_{22}$.

It is easy to see that H exists if and only if

$$MQ = P$$

for some vector $Q \geq 0$ (componentwise) of probabilities. The vectors P for which such a Q exists are exactly those within the polytope whose

Table 1: Characterizations of the sets of four different types (columns) subject to three constrains (rows).

	All (p,ε)	Fit (ε)	Force (ε)	Equi (ε)
chaos	$\in$ ELFP	$\max \mathsf{S}\varepsilon \leq {}^1\!/_2$	$\in \left[0, \frac{1}{2}\right]^4$	$\max \mathsf{S}\varepsilon \leq {}^1\!/_2$
quant	$\max \mathsf{S}_1 p \leq {}^{\sqrt{2}}\!/_2$	$\max \mathsf{S}_0\varepsilon \leq \frac{3-\sqrt{2}}{2}$, $\max \mathsf{S}_1\varepsilon \leq {}^1\!/_2$	$\max \mathsf{S}_0\varepsilon \geq \frac{3-\sqrt{2}}{2}$	$\frac{3-\sqrt{2}}{2} \in \mathsf{S}_0\varepsilon$, $\max \mathsf{S}_1\varepsilon \leq {}^1\!/_2$
class	$\max \mathsf{S}_1 p \leq {}^1\!/_2$	$\max \mathsf{S}_1\varepsilon \leq {}^1\!/_2$	$1 \in \mathsf{S}_0\varepsilon$	$1 \in \mathsf{S}_0\varepsilon$

vertices are the columns of the matrix M. The term ELFP is due to this construction extending that of the linear feasibility test in [*11*]. This test, among other applications, is the most general way of extending the Bell/CHSH criterion to an arbitrary number of particles, spin axes, and spin quantum numbers [*11,12,29-31*].

To describe ELFP by inequalities on (p,ε), we introduce the 16-component sets

$$\mathsf{S}p = \left\{ \begin{array}{l} \pm\left(p_{11} - {}^1\!/_4\right) \pm \left(p_{12} - {}^1\!/_4\right) \\ \pm\left(p_{21} - {}^1\!/_4\right) \pm \left(p_{22} - {}^1\!/_4\right) : \\ \text{each } \pm \text{ is } + \text{ or } - \end{array} \right\},$$

$$\mathsf{S}\varepsilon = \left\{ \begin{array}{l} \pm\left(\varepsilon_1^1 - {}^1\!/_4\right) \pm \left(\varepsilon_1^2 - {}^1\!/_4\right) \\ \pm\left(\varepsilon_2^1 - {}^1\!/_4\right) \pm \left(\varepsilon_2^2 - {}^1\!/_4\right) : \\ \text{each } \pm \text{ is } + \text{ or } - \end{array} \right\}.$$

$\mathsf{S}_0 p$ and $\mathsf{S}_1 p$ denote the subsets of $\mathsf{S}p$ with, respectively, even (0,2, or 4) and odd (1 or 3) number of + signs; $\mathsf{S}_0\varepsilon$ and $\mathsf{S}_1\varepsilon$ are defined analogously. ELFP is described by

$$\max\left(\max \mathsf{S}_0 p + \max \mathsf{S}_1\varepsilon,\ \max \mathsf{S}_1 p + \max \mathsf{S}_0\varepsilon\right) \leq {}^3\!/_2 \tag{4.1}$$

(see Appendix B1).

5. ALL-FIT-FORCE-EQUI SETS AND CONSTRAINTS ON OBSERVABLES

Let constr (p) denote any constraint (e.g., inequalities) imposed on p. Our approach consists in characterizing this constraint by solving the following four problems:

1. Find the set $\text{All}_{\text{constr}}$ of all $(p,\varepsilon) \in [0, {}^1\!/_2]^8$ with p subject to constr (p):

$$(p,\varepsilon) \in \text{All}_{\text{constr}} \Longleftrightarrow (\text{constr}(p) \text{ and } (p,\varepsilon) \in \text{ELFP}).$$

2. Find the set $\text{Fit}_{\text{constr}}$ of connection vectors $\varepsilon \in [0, {}^1\!/_2]^4$ that fit (are compatible with) all observables satisfying constr:

$$\varepsilon \in \text{Fit}_{\text{const}} \Longleftrightarrow (\text{constr}(p) \Longrightarrow (p,\varepsilon) \in \text{ELFP}).$$

3. Find the set $\text{Force}_{\text{constr}}$ of $\varepsilon \in [0, {}^1\!/_2]^4$ that force all observables p to satisfy constr:

$$\varepsilon \in \text{Force}_{\text{constr}} \Longleftrightarrow ((p,\varepsilon) \in \text{ELFP} \Longrightarrow \text{constr}(p)).$$

4. Find the set $\text{Equi}_{\text{constr}}$ of $\varepsilon \in [0, {}^1\!/_2]^4$ for which an observable p satisfies constr if and only if (p,ε) is in the ELFP set:

$$\varepsilon \in \text{Equi}_{\text{constr}} \Longleftrightarrow (\text{constr}(p) \Longleftrightarrow (p,\varepsilon) \in \text{ELFP}).$$

Clearly, $\text{Equi}_{\text{constr}} = \text{Force}_{\text{constr}} \cap \text{Fit}_{\text{constr}}$.

To illustrate, we focus on the following four benchmark constraints. The no-constraint situation is given by

$$\text{chaos(p)} \Longleftrightarrow p \in [0, {}^1\!/_2]^4,$$

equivalent to (2.1) with $C = 2$. The quantum constraint is given by

$$\text{quant}(p) \Longleftrightarrow \max \mathsf{S}_1 p \leq {}^{\sqrt{2}}\!/_2,$$

equivalent to (2.1) with $C = \sqrt{2}$. The "classical" constraints are given by

$$\text{class}(p) \Longleftrightarrow \max \mathsf{S}_1 p \leq {}^1\!/_2,$$

equivalent to the Bell/CHSH inequalities ($C = 1$). Finally, we consider the constraint

$$\text{fix}\,(p) \Longleftrightarrow p = \text{specific vector}.$$

For all constraints except for fix (p) the sets All, Fit, Force, and Equi are as shown in Table 1 (for derivations see Appendix B2).

Thus, $\text{Fit}_{\text{chaos}}$ is the set of all ε such that $\max \mathsf{S}\varepsilon \leq {}^1\!/_2$: if an ε is in this set, then any p (no constraints) is compatible with it. $\text{Force}_{\text{quant}}$ is characterized by $\max \mathsf{S}_0\varepsilon \geq \frac{3-\sqrt{2}}{2}$: if an ε is in this set, then all compatible with it p satisfy quant(p). $\text{Equi}_{\text{class}}$ is the set of all ε such that $\mathsf{S}_0\varepsilon$ contains 1: for any such an ε, a p is compatible with it if and only if it satisfies class (p).

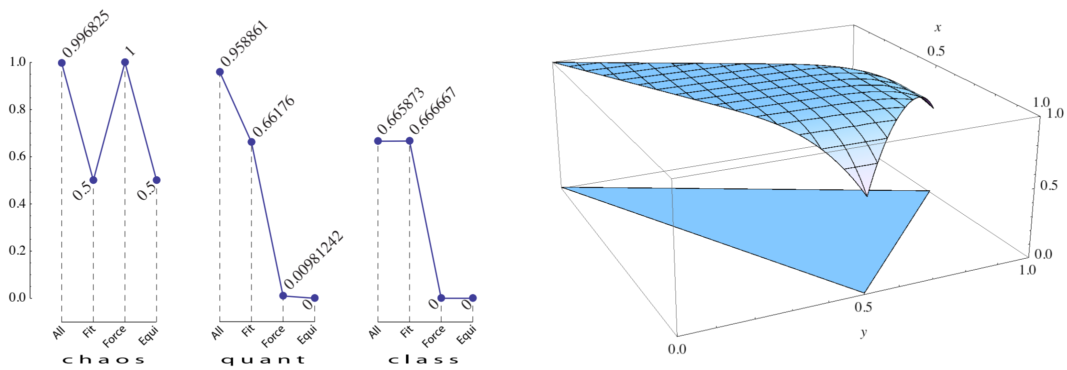

Figure 1: Left: Profiles $\mathrm{Vol}^8(\mathrm{All}_{\mathrm{constr}}) \rightarrow \mathrm{Vol}^4(\mathrm{Fit}_{\mathrm{constr}}) \rightarrow \mathrm{Vol}^4(\mathrm{Force}_{\mathrm{constr}}) \rightarrow \mathrm{Vol}^4(\mathrm{Equi}_{\mathrm{constr}})$ for constraints chaos, quant, and class. Right: $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{fix}(p)}\right)$ as a function of $x = \max \mathsf{S}_0 p$ and $y = \max \mathsf{S}_1 p$. The possible (x, y)-pairs form the triangle $((0,0), (1/2, 1), (1, 1/2))$, and $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{fix}(p)}\right) = 1 + \frac{\rho(x)}{3}\left(-1 + 8x - 24x^2 + 32x^3 - 16x^4\right) + \frac{\rho(y)}{3}\left(-1 + 8y - 24y^2 + 32y^3 - 16y^4\right)$, where $\rho(z) = 1$ if $z \geq 1/2$ and $\rho(z) = 0$ otherwise.

For each of these sets we compute Vol^d, its volume normalized by that of $[0, 1/2]^d$, with d being the dimensionality of the set (Fig. 1, left). Thus, the defining property of $\mathrm{Force}_{\mathrm{class}}$, $1 \in \mathsf{S}_0\varepsilon$, is satisfied if and only if either all ε_i^k are 0, or they all are $1/2$, or two of them are 0 and two $1/2$. Hence $\mathrm{Vol}^4(\mathrm{Force}_{\mathrm{class}}) = 0$. For nonzero volumes, the derivation is described in Appendix B2. Each panel of Fig. 1, left, can be viewed as a "profile" of the corresponding constraint. Each of the first three volumes in a panel can be viewed as characterizing the "strictness" of constraint, in three different meanings. The intuition of a stricter constraint is that it corresponds to a smaller $\mathrm{Vol}^8(\mathrm{All}_{\mathrm{constr}})$, larger $\mathrm{Vol}^4(\mathrm{Fit}_{\mathrm{constr}})$, and smaller $\mathrm{Vol}^4(\mathrm{Force}_{\mathrm{constr}})$.

The constraint $\mathrm{fix}(p)$ has to be handled separately. Clearly, $\mathrm{Vol}^8\left(\mathrm{All}_{\mathrm{fix(p)}}\right) = 0$. $\mathrm{Fit}_{\mathrm{fix}(p)}$ is described by

$$\begin{array}{l} \max \mathsf{S}_1\varepsilon \leq 3/2 - \max \mathsf{S}_0 p, \\ \max \mathsf{S}_0\varepsilon \leq 3/2 - \max \mathsf{S}_1 p, \end{array} \tag{5.1}$$

and $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{fix}(p)}\right)$ is a polynomial function of $\max \mathsf{S}_0 p$ and $\max \mathsf{S}_1 p$, these two quantities forming the triangle $((0,0), (1/2, 1), (1, 1/2))$. The polynomial and its values are shown in Fig. 1, right (see Appendix B3, for computational details). $\mathrm{Force}_{\mathrm{fix}(p)}$ is clearly empty, hence so is $\mathrm{Equi}_{\mathrm{fix}(p)}$.

6. CONCLUSION

At the cost of greater computational complexity but with no conceptual complications the computations involved in this demonstration of the all-possible-couplings approach to probabilistic contextuality can be extended to more general cases: arbitrary marginal probabilities, nonlinear constraints, and greater numbers of inputs, outputs, and their possible values. The language for a completely general theory, involving unrestricted (not necessarily finite) sets of inputs, outputs, and their values, is presented in Appendix A.

Acknowledgments

Supported by NSF grant SES-1155956. We thank Joseph Malpeli and Lasse Leskelä for helpful discussions, and Damir Dzhafarov for suggestions on notation and terminology.

APPENDIX A: All-possible-couplings approach on the general level

We show here how the approach presented in the main text generalizes to arbitrary sets of inputs and random outputs. We use the term *sequence* to refer to any indexed family (a function from an index set into a set), with index sets not necessarily countable. We present se-

quences in the form $(x^y : y \in Y)$, $(x_z : z \in Z)$, or $(x_z^y : y \in Y, z \in Z)$. A random variable is understood most broadly, as a measurable mapping between any two probability spaces. In particular, any sequence of jointly distributed random variables is a random variable. For brevity, we omit an explicit presentation of probability spaces and distributions. In all other respects the notation and terminology closely follow [*15,12*].

An *input* is a set of elements called *input values*. Let $\alpha = \left(\alpha^k : k \in K\right)$ be a sequence of inputs. A *treatment* is a sequence $\phi = (x^k : k \in K)$ that belongs to a nonempty set $\Phi \subset \prod_{k\in K} \alpha^k$ (so that $x^k \in \alpha^k$ for all $k \in K$). If $\phi \in \Phi$, $k \in K$, and $I \subset K$, then $\phi(k) = x^k \in \alpha^k$ and $\phi|I$ is the restriction of ϕ to I, i.e., the sequence $(x^k : k \in I)$.

An *output* is a random variable. Let $\left(A_\phi^k : k \in K, \phi \in \Phi\right)$ be a sequence of outputs such that

1. $A_\phi = \left(A_\phi^k : k \in K\right)$ is a random variable for every $\phi \in \Phi$, i.e., the random variables A_ϕ^k across all possible k possess a joint distribution;

2. if $\phi, \phi' \in \Phi$, $I \subset K$, and $\phi|I = \phi'|I$, then $\left(A_\phi^k : k \in I\right) \sim \left(A_{\phi'}^k : k \in I\right)$.

Property 2 is *(complete) marginal selectivity* [*9*]. A_ϕ is called an *observable* (random variable), and $A = (A_\phi : \phi \in \Phi)$ is the *sequence of observables.*

Remark A.1. The interpretation is that for every ϕ, each α^k may "directly" influence A_ϕ^k but no other output in A_ϕ. The fact that inputs in $\alpha = \left(\alpha^k : k \in K\right)$ and outputs in an observable $A_\phi = \left(A_\phi^k : k \in K\right)$ are in a bijective correspondence is not restrictive: this can always be achieved by an appropriate grouping of inputs and (re)definition of treatments ϕ [*11*].

Remark A.2. The case considered in the main text corresponds to $K = \{1,2\}$,

$$\alpha = \left(\alpha^1, \alpha^2\right) \text{ with } \alpha^k = \left\{\alpha_1^k, \alpha_2^k\right\} \text{ for } k \in \{1,2\},$$

$$\begin{aligned}&\Phi = \{\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}\}\\ &\text{with } \phi_{ij} = \left(\alpha_i^1, \alpha_j^2\right) \text{ for } i,j \in \{1,2\},\end{aligned}$$

and (abbreviating $A_{\phi_{ij}}$ as A_{ij} and $A_{\phi_{ij}}^k$ as A_{ij}^k)

$$\begin{aligned}&A = (A_{11}, A_{12}, A_{21}, A_{22}),\\ &\text{with } A_{ij} = \left(A_i^1, A_j^2\right) \text{ for } i,j \in \{1,2\},\end{aligned}$$

where each A_{ij}^k is a binary random variable with $\Pr\left[A_{ij}^k = a_1^k\right] = \Pr\left[A_{ij}^k = a_2^k\right] = 1/2$.

Given a sequence of observables $A = (A_\phi : \phi \in \Phi)$, a sequence of random variables

$$C_A = \left(C_\tau^I : \tau \in \prod_{k\in I} \alpha^k, I \in 2^K - \{\emptyset, K\}\right)$$

(not necessarily jointly distributed) is called a *connecting set* for A if each C_τ^I is a coupling for

$$A_\tau^I = \left(A_\phi^I : \phi \in \Phi, \phi|I = \tau\right),$$

where $A_\phi^I = \left(A_\phi^k : k \in I\right)$. This means that C_τ^I is a random variable of the form

$$C_\tau^I = \left(C_{\tau,\phi}^I : \phi \in \Phi, \phi|I = \tau\right)$$

with

$$C_{\tau,\phi}^I \sim A_\phi^I$$

for all $\phi \in \Phi$ such that $\phi|I = \tau$. C_τ^I is called an (I,τ)-*connection.* The indexation in $C_{\tau,\phi}^I$ is to ensure that if $(I,\tau) \neq (I',\tau')$, then C_τ^I and $C_{\tau'}^{I'}$ are stochastically unrelated. An *identity* (I,τ)-*connection* C_τ^I is one with $\Pr\left[C_{\tau,\phi}^I = C_{\tau,\phi'}^I\right] = 1$ for any $\phi, \phi' \in \Phi$. Formally, A itself can be viewed as a connection $C_\emptyset^\emptyset$, but it is preferable to keep A separate by not allowing $I = \emptyset$, as A is the only observable part of the construction.

Remark A.3. It is generally convenient not to distinguish identically distributed connections. By abuse of language, the distribution of C_τ^I (or some characterization thereof) can also be called (I,τ)-connection. We use this language in the main text when we represent $\left(\{k\}, k \mapsto \alpha_i^k\right)$-connections (without introducing them explicitly) by probabilities ε_i^k and call ε a connection vector. See Remark A.4.

A jointly distributed sequence

$$H = \left(H_\phi^k : k \in K, \phi \in \Phi\right)$$

is called an *Extended Joint Distribution Sequence* (EJDS) for (A, C_A) if for any $I \in 2^K - \{\emptyset, K\}$ and any $\tau \in \prod_{k\in I} \alpha^k$,

$$H_\tau^I = \left(H_\phi^I : \phi \in \Phi, \phi|I = \tau\right) \sim C_\tau^I,$$

where $H^I_\phi = \left(H^k_\phi : k \in I\right)$, and

$$H^K_\phi = \left(H^k_\phi : k \in K\right) \sim A_\phi$$

for any $\phi \in \Phi$.

Remark A.4. For the case considered in the main text, a connecting set for A is (conveniently replacing $C^{\{k\}}_{\phi_{ij}}$, $C^{\{1\}}_{\phi_{ij}|\{1\}}$, and $C^{\{2\}}_{\phi_{ij}|\{2\}}$ with C^k_{ij} , C^1_i, and C^2_j, respectively)

$$\begin{array}{l} C_A = \left(C^1_1, C^1_2, C^2_1, C^2_2\right) \\ \text{with } C^1_i = \left(C^1_{i,i1}, C^1_{i,i2}\right) \text{ and } C^2_j = \left(C^2_{j,1j}, C^2_{j,2j}\right), \end{array}$$

such that

$$C^1_{i,ij} \sim A^1_{ij},\ C^2_{j,ij} \sim A^2_{ij} \text{ for } i,j \in \{1,2\}\,.$$

An EJDS for (A, C_A) is a random variable (using analogous abbreviations)

$$H = \left(H^1_{11}, H^2_{11}, H^1_{12}, H^2_{12}, H^1_{21}, H^2_{21}, H^1_{22}, H^2_{22}\right)$$

such that

$$\left(H^1_{i1}, H^1_{i2}\right) \sim C^1_i,\ \left(H^2_{1j}, H^2_{2j}\right) \sim C^2_j$$

and

$$H^{12}_{ij} = \left(H^1_{ij}, H^2_{ij}\right) \sim A_{ij} = \left(A^1_{ij}, A^2_{ij}\right) \text{ for } i,j \in \{1,2\}$$

In the main text each C^k_i is represented by ε^k_i and each H^{12}_{ij} by p_{ij}.

An EJDS for (A, C_A) reduces to the Joint Distribution Criterion set (JDC set) of the theory of selective influences [*14-12*] if all connections in C_A are identity ones. Note that no connection has an empirical meaning: for distinct $\phi, \phi' \in \Phi$, the variables A^I_ϕ and $A^I_{\phi'}$ corresponding to $C^I_{\tau,\phi}$ and $C^I_{\tau,\phi'}$ do not have an observable (or theoretically privileged) pairing scheme.

Let X be any set whose elements are sequences of observables $A = (A_\phi : \phi \in \Phi)$. X can be viewed as the set of all possible observables satisfying certain constraints. We define the sets All_X, Fit_X, Force_X, and Equi_X as follows:

1. All_X is the set of all pairs (A, C_A) such that

$$A \in X \text{ and there exists an EJDS } H \text{ for } (A, C_A)\,.$$

2. Fit_X is the set of all C_A such that

$$A \in X \Longrightarrow \text{there exists an EJDS } H \text{ for } (A, C_A)\,.$$

3. Force_X is the set of all C_A such that

$$\text{there exists an EJDS } H \text{ for } (A, C_A) \Longrightarrow A \in X.$$

4. $\text{Equi}_X = \text{Force}_X \cap \text{Fit}_X$, that is, $C_A \in \text{Equi}_X$ if and only if

$$A \in X \Longleftrightarrow \text{there exists an EJDS } H \text{ for } (A, C_A)\,.$$

The all-possible-couplings approach in the general case consists in characterizing any X (interpreted as a type of contextuality or determinism) by All_X, Fit_X, Force_X, and Equi_X. A straightforward generalization of this approach that might be useful in some applications is to replace C_A in all definitions with a subset of C_A, or several subsets of C_A tried in turn. Thus one might consider connections involving only particular $I \subset K$ (e.g., only singletons), or one might require that some of the connections are identity ones.

APPENDIX B: Computational Details

B1. Computations for ELFP

A convex bounded polytope can be equivalently defined either as the convex hull of a set of points (V-representation) or as the intersection of half-spaces (H-representation). For our purposes, a V-representation of a convex polytope in d-space is given by a set of points $x_1, \ldots, x_n \in \mathbb{R}^d$. The polytope consists of all convex combinations of these points: $\lambda_1 x_1 + \cdots + \lambda_n x_n$, for all $\lambda_1, \ldots, \lambda_n \geq 0$, $\lambda_1 + \cdots + \lambda_n = 1$. It is possible that the polytope is of lower dimension than the space $\mathbb{R}^d$ in which it is defined if all the points x_i reside in a lower dimensional affine subspace of $\mathbb{R}^d$. A minimal V-representation (including only extreme points, i.e., points that are vertices of the polytope) is unique. The H-representation of a convex polytope is given by vectors $a_1, \ldots, a_m \in \mathbb{R}^d$ and a vector $b \in \mathbb{R}^m$. The polytope consists of the points $x \in \mathbb{R}^d$ satisfying $a_i^T x \leq b_i$ for all $i = 1, \ldots, m$. A lower-dimensional convex polytope can be represented by including inequalities of the forms $a^T x \leq b$ and $(-a)^T x \leq -b$ for some a and b or by explicitly specifying certain constraints as equations in the representation. For a full-dimensional convex polytope, the minimal H-representation is unique. However, for a lower-dimensional polytope, the equation constraints can be specified in many equivalent ways and the

inequality constraints can look different depending on which of the linearly related coordinates are used to specify them.

There exist algorithms for converting between the two representations of a convex polytope in exact rational arithmetic. We have used our own program for these conversions but other programs, such as *lrs* (http://cgm.cs.mcgill.ca/~avis/C/lrs.html), can do the same. The conversion between the two representation is computationally demanding, the algorithms generally requiring superpolynomial time in the size of the input.

A computationally simpler problem is eliminating redundant points (those that are not vertices of the polytope) from a V-representation or eliminating redundant equations or inequalities from an H-representation. This problem can be solved by linear programming and the algorithm is implemented in the *redund* program that comes with *lrs*. However, the *redund* program is not sufficient for putting an H-representation to a minimal form as it cannot convert sets of inequalities into equivalent equations (e.g., the three inequalities $x \geq 0$, $y \geq 0$, $x + y \leq 0$ should be minimally represented as the two equations $x = 0$, $y = 0$). To find the minimal H-representation, for every constraint $a_i^T x \leq b_i$ or $a_i^T x = b_i$ in turn, one can find the upper and lower bounds u and l by maximizing and minimizing the expression $a_i^T x$ given the other constraints, and apply the following rules:

1. if this is an equation constraint (i.e., $a_i^T x = b_i$) and $u = l = b_i$, then the constraint is redundant and can be eliminated;

2. if this is an inequality constraint (i.e., $a_i^T x \leq b_i$) and $u \leq b_i$, then the inequality is redundant and can be eliminated. Otherwise, if $l = b_i$, then the constraint should be converted to an equation.

The dimension of a polytope can be determined from a minimal H-representation. It is the dimension of the space minus the number of equation constraints in the minimal representation. Given a full-dimensional polytope, its volume can be computed using the *lrs* program alongside the conversion from a V-representation to an H-representation. If the polytope is given as an H-representation, then it has to be converted to a V-representation first to compute its volume using *lrs*. To compute the volume of a lower-dimensional polytope, we first move to a lower-dimensional parameterization that spans the affine subspace where the polytope resides.

To compute ELFP, we begin by formulating the linear programming problem $MQ = P$ subject to $Q \geq 0$, as described in the main text (M being $2^5 \times 2^8$, P having 2^5 components). M defines the V-representation for ELFP, and Vol^8 for ELFP is computed directly from it. Applying an algorithm to find an equivalent H-representation we obtain a system of 160 inequalities and 16 equations. We can then substitute the expressions in the above matrices into this system and reduce any redundant inequalities and equations. The resulting system has 144 nonredundant inequalities and no equations with the $p_{11}, p_{12}, p_{21}, p_{22}, \varepsilon_1^1, \varepsilon_2^1, \varepsilon_1^2, \varepsilon_2^2$ variables. Then we algebraically simplify the list of 144 inequalities, first into

$$\begin{aligned}
-\Gamma \leq -p_{11} + p_{21} + p_{12} + p_{22} \leq 1 + \Gamma,\\
-\Gamma \leq p_{11} - p_{21} + p_{12} + p_{22} \leq 1 + \Gamma,\\
-\Gamma \leq p_{11} + p_{21} - p_{12} + p_{22} \leq 1 + \Gamma,\\
-\Gamma \leq p_{11} + p_{21} + p_{12} - p_{22} \leq 1 + \Gamma,
\end{aligned}$$

$$-\Lambda \leq p_{11} + p_{21} + p_{12} + p_{22} \leq 2 + \Lambda,$$

$$\begin{aligned}
|-p_{11} - p_{21} + p_{12} + p_{22}| \leq 1 + \Lambda,\\
|-p_{11} + p_{21} - p_{12} + p_{22}| \leq 1 + \Lambda,\\
|-p_{11} + p_{21} + p_{12} - p_{22}| \leq 1 + \Lambda,
\end{aligned}$$

where

$$\begin{aligned}
\Gamma = \min\{\, & 1 - \varepsilon_1^1 - \varepsilon_1^2 + \varepsilon_2^1 + \varepsilon_2^2,\\
& 1 - \varepsilon_1^1 + \varepsilon_1^2 - \varepsilon_2^1 + \varepsilon_2^2,\\
& 1 - \varepsilon_1^1 + \varepsilon_1^2 + \varepsilon_2^1 - \varepsilon_2^2,\\
& 1 + \varepsilon_1^1 - \varepsilon_1^2 - \varepsilon_2^1 + \varepsilon_2^2,\\
& 1 + \varepsilon_1^1 - \varepsilon_1^2 + \varepsilon_2^1 - \varepsilon_2^2,\\
& 1 + \varepsilon_1^1 + \varepsilon_1^2 - \varepsilon_2^1 - \varepsilon_2^2,\\
& \varepsilon_1^1 + \varepsilon_1^2 + \varepsilon_2^1 + \varepsilon_2^2,\\
& 2 - \varepsilon_1^1 - \varepsilon_1^2 - \varepsilon_2^1 - \varepsilon_2^2 \,\},
\end{aligned}$$

$$\begin{aligned}\Lambda = \min\{ & -\varepsilon_1^1+\varepsilon_1^2+\varepsilon_2^1+\varepsilon_2^2,\\ & \varepsilon_1^1-\varepsilon_1^2+\varepsilon_2^1+\varepsilon_2^2,\\ & \varepsilon_1^1+\varepsilon_1^2-\varepsilon_2^1+\varepsilon_2^2,\\ & \varepsilon_1^1+\varepsilon_1^2+\varepsilon_2^1-\varepsilon_2^2,\\ & 1-\varepsilon_1^1-\varepsilon_1^2-\varepsilon_2^1+\varepsilon_2^2,\\ & 1-\varepsilon_1^1-\varepsilon_1^2+\varepsilon_2^1-\varepsilon_2^2,\\ & 1-\varepsilon_1^1+\varepsilon_1^2-\varepsilon_2^1-\varepsilon_2^2,\\ & 1+\varepsilon_1^1-\varepsilon_1^2-\varepsilon_2^1-\varepsilon_2^2\},\end{aligned}$$

and then, by noticing regularities, into the compact inequality (4.1).

Remark. Changing $\varepsilon_j^i \to 1/2-\varepsilon_j^i$ leads to (denoting the new ε-vector by ε')

$$\max \mathsf{S}_1\varepsilon' = \max \mathsf{S}_0\varepsilon, \max \mathsf{S}_0\varepsilon' = \max \mathsf{S}_1\varepsilon.$$

Analogously for $p_{ij} \to 1/2 - p_{ij}$,

$$\max \mathsf{S}_1 p' = \max \mathsf{S}_0 p, \max \mathsf{S}_0 p' = \max \mathsf{S}_1 p.$$

It follows that we cannot without loss of generality confine all components of ε or p to $[0, 1/4]$. But ELFP does not change if the transformation $x \to 1/2 - x$ is applied to an even number of the components of (p, ε).

B2. Computations for chaos(p), quant(p), and class(p) constraints

The $\mathrm{All}_{\mathrm{constr}}$ polytopes for the three constraints are obtained by concatenating the ELFP equations and inequalities with the constraint inequalities. Then, the volumes are computed by using the *lrs* program as described above.

For $\mathrm{Fit}_{\mathrm{constr}}$ polytopes, we observe first that they are convex. This follows from

$$\begin{aligned}\mathrm{Fit}_{\mathrm{constr}} &= \{\varepsilon : \forall i = 1,\dots,n : \left(p_{(i)},\varepsilon\right) \in \mathrm{ELFP}\,\}\\ &= \mathrm{ELFP}_{p_{(1)}} \cap \dots \cap \mathrm{ELFP}_{p_{(n)}},\end{aligned}$$

where $p_{(i)}$, $i = 1,\dots,n$, denote the vertices of the 4D convex polytope defined by constr and $\mathrm{ELFP}_{p_{(i)}}$ denotes the (convex) cross-section of the ELFP set formed with $p = p_{(i)}$. It follows that $\mathrm{Fit}_{\mathrm{constr}}$ is convex as the intersection of convex sets. Following the logic of this observation, we have implemented a general program for eliminating variables from a system of linear equations and inequalities so that the resulting system is satisfied for exactly those values for which there exist such values of the eliminated variables for which the original system is satisfied. This program together with steps to ensure that the resulting representation is minimal was used to find all the Fit sets shown in the main text.

Finding the forcing sets is more difficult as they are generally not convex. We characterize them using the equation

$$\begin{aligned}&\mathrm{Force}_{\mathrm{chaos}} - \mathrm{Force}_{\mathrm{constr}}\\ &= \{\varepsilon : (\exists p : (p,\varepsilon) \in \mathrm{ELFP} \wedge \neg \mathrm{constr}\,(p))\}\,.\end{aligned}$$

This equation provides an algorithm: for each inequality in constr, form the conjunction of the ELFP inequalities with the negation of the inequality. Then project this conjunction to the ε 4-space. The union of these projections over all inequalities in constr is the set $\mathrm{Force}_{\mathrm{chaos}} - \mathrm{Force}_{\mathrm{constr}}$. We have implemented a general program that takes as input a representation of a polytope, a list of additional constraints, and a list of variables to eliminate. It then outputs a representation of the difference of the polytope and the set represented by the additional constraints projected to the remaining (not eliminated) variables. This representation consists of a list of linear systems whose disjunction characterizes the resulting set. In all our computations it turned out that all the linear systems in the disjunction were the same, and so the sets $\mathrm{Force}_{\mathrm{chaos}} - \mathrm{Force}_{\mathrm{constr}}$ are in fact convex in these cases.

The computations of Equi sets require no elaboration.

Remark. There is the practical problem that the negation of a $\leq$ -inequality is a $>$ -inequality while standard algorithms only accept closed convex polytopes. To cope with this problem, we approximated $a > b$ by $a \geq b+$(very small number). We also used a rational approximation to $\sqrt{2}$ in the quant constraints. In both cases, we have repeated the computations with decreasing values of "very small number" until it was obvious where the results converged.

B3. Computations for $\mathrm{Fit}_{\mathrm{fix}(p)}$ constraint

That $\max \mathsf{S}_0 p$ and $\max \mathsf{S}_1 p$ are contained in and completely fill the triangle $\{(0,0),(1/2,1),(1,1/2)\}$ can be verified by splitting (5.1) into 64 component cases according as which of the values of $\mathsf{S}_0 p$ and $\mathsf{S}_1 p$ are the

maxima, finding the vertices of each component system, and drawing the union of these components in $\max \mathsf{S}_0 p$ and $\max \mathsf{S}_1 p$ coordinates. The triangle is described by

$$\begin{aligned} 2\max \mathsf{S}_0 p - \max \mathsf{S}_1 p &\geq 0, \\ 2\max \mathsf{S}_1 p - \max \mathsf{S}_0 p &\geq 0, \\ \max \mathsf{S}_0 p + \max \mathsf{S}_1 p &\leq {}^3\!/_2. \end{aligned}$$

Adding these inequalities to the representation of (5.1) as linear inequalities according to the definitions of $\max \mathsf{S}_0 \varepsilon$ and $\max \mathsf{S}_1 \varepsilon$, we obtain a 6D polytope $P^{(6)}$ in $(\varepsilon, \max \mathsf{S}_0 p, \max \mathsf{S}_1 p)$-coordinates. In the V-representation of $P^{(6)}$, all vertices have values of $\max \mathsf{S}_0 p$ and $\max \mathsf{S}_1 p$ in the set

$$\{(0,0), ({}^1\!/_4, {}^1\!/_2), ({}^1\!/_2, {}^1\!/_4), ({}^1\!/_2, 1), (1, {}^1\!/_2)\}.$$

It follows that every edge of the polytope projects to one of these 5 points or to a line connecting two of them. Consequently, as $(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon)$ changes within any triangle T formed by these lines, the cross-section $P^{(4)}_{(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon)}$ of $P^{(6)}$ retains its structure (face lattice) while its coordinates change as affine functions of $(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon) \in T$. It follows that the volume of $P^{(4)}_{(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon)}$ is a polynomial of $(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon) \in T$ of at most degree four. The coefficients of these polynomials were obtained by fitting unconstrained degree 4 polynomials to the exact volumes $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{fix}(p)}\right)$ for $(\max \mathsf{S}_0 p, \max \mathsf{S}_1 p) \in \{0, .01, .02, \ldots, 1\}^2$. It turns out that the coefficients change only if either of the differences $\max \mathsf{S}_0 p - {}^1\!/_2$ and $\max \mathsf{S}_1 p - {}^1\!/_2$ changes its sign. In all cases the fit is perfect for the number of points far exceeding the number of coefficients, confirming that the computations are correct.

REFERENCES

1. Townsend JT, Schweickert R (1989) Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *J. Math. Psych.* **33**: 309-327.
2. Cereceda J (2000) Quantum mechanical probabilities and general probabilistic constraints for Einstein–Podolsky–Rosen–Bohm experiments. *Found. Phys. Lett.* **13**: 427-442
3. Bell J (1964) On the Einstein-Podolsky-Rosen paradox. *Physics* **1**: 195-200.
4. Bell J (1966) On the problem of hidden variables in quantum mechanics. *Rev. Modern Phys.* **38**: 447-453.
5. Clauser JF, Horne MA, Shimony A, Holt RA (1969) Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**: 880-884.
6. Clauser JF, Horne MA (1974) Experimental consequences of objective local theories. *Phys. Rev. D*, **10**: 526-535.
7. Fine A (1982) Joint distributions, quantum correlations, and commuting observables. *J. Math. Phys.* **23**: 1306-1310.
8. Fine A (1982) Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.* **48**: 291-295.
9. Dzhafarov EN (2003) Selective influence through conditional independence. *Psychom.* **68**: 7-26.
10. Bohm D, Aharonov Y (1957) Discussion of experimental proof for the paradox of Einstein, Rosen and Podolski. *Phys. Rev.*, **108**: 1070-1076.
11. Dzhafarov EN, Kujala JV (2012) Selectivity in probabilistic causality: Where psychology runs into quantum physics. *J. Math. Psych.*, **56**: 54-63.
12. Dzhafarov EN, Kujala JV (in press) Quantum entanglement and the issue of selective influences in psychology: An overview. *Lect. Notes in Comp. Sci.* (available as arXiv:1209.0041v1 [quant-ph])
13. Kujala JV, Dzhafarov EN (2008) Testing for selectivity in the dependence of random variables on external factors. *J. Math. Psych.* **52**: 128-144.
14. Dzhafarov EN, Kujala JV (2010) The Joint Distribution Criterion and the Distance Tests for selective probabilistic causality. *Front. Quant. Psych. Meas.* **1:151** doi: 10.3389/fpsyg.2010.00151.
15. Dzhafarov EN, Kujala JV (in press) Order-distance and other metric-like functions on jointly distributed random variables.

Proc. Amer. Math. Soc. (available as arXiv:1110.1228v3 [math.PR])

16. Dzhafarov EN, Schweickert R, Sung K (2004) Mental architectures with selectively influenced but stochastically interdependent components. *J. Math. Psych.*, **48**: 51-64.

17. Schweickert R, Fisher DL, Goldstein WM (2010) Additive factors and stages of mental processes in task networks. *J. Math. Psych.* **54**: 405-414.

18. Sternberg S (1969) The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), *Attention and Performance II. Acta Psych.*, **30**: 276–315.

19. Townsend JT (1984) Uncovering mental processes with factorial experiments. *J. Math. Psych.*, **28**: 363–400.

20. Suppes P, Zanotti M (1981) When are probabilistic explanations possible? *Synthese* **48**: 191-199.

21. Suppes P (2002) Representation *and Invariance of Scientific Structures* (CSLI, Stanford) pp. 332-351.

22. Aspect A, Grangier P, Roger G (1981) Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.* **47**: 460–463.

23. Aspect A, Grangier P, Roger G (1982) Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: A new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**: 91–94.

24. Weihs G, Jennewein T, Simon C, Weinfurter H., Zeilinger A. (1998) Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* **81**: 5039–5043.

25. Thorisson H (2000) *Coupling, Stationarity, and Regeneration* (New York, Springer).

26. Laudisa F (1997) Contextualism and nonlocality in the algebra of EPR observables. *Phil. Sci.*, **64**: 478-496.
27. Khrennikov AYu (2008) EPR–Bohm experiment and Bell's inequality: Quantum physics meets probability theory. *Theor. Math. Phys.*, **157**: 1448–1460.

28. Khrennikov AYu (2009) *Contextual Approach to Quantum Formalism. Fundamental Theories of Physics* **160** (Dordrecht, Springer).

29. Pitowsky I *(*1989) *Quantum Probability-Quantum Logic, Lect. Notes in Phys.* **321** (Heidelberg, Springer).

30. Peres A (1999) All the Bell inequalities. *Found. Phys.* **29** :589-614.

31. Basoalto RM, Percival IC (2003) BellTest and CHSH experiments with more than two settings. *J. Phys. A: Math. & Gen.* **36**: 7411–7423.